

⑨ 日本国特許庁(JP)

⑩ 特許出願公開

⑫ 公開特許公報(A) 平4-52741

⑮ Int. Cl.⁵

G 06 F 12/08
12/12

識別記号

M
A

庁内整理番号

7232-5B
7232-5B

⑭ 公開 平成4年(1992)2月20日

審査請求 未請求 請求項の数 7 (全10頁)

⑮ 発明の名称 キャッシュメモリ装置

⑯ 特 願 平2-155776

⑰ 出 願 平2(1990)6月14日

⑱ 発 明 者 相 川 健 神奈川県川崎市幸区小向東芝町1番地 株式会社東芝総合
研究所内

⑲ 発 明 者 皆 川 健 二 神奈川県川崎市幸区小向東芝町1番地 株式会社東芝総合
研究所内

⑳ 発 明 者 斉 藤 光 男 神奈川県川崎市幸区小向東芝町1番地 株式会社東芝総合
研究所内

㉑ 出 願 人 株 式 会 社 東 芝 神奈川県川崎市幸区堀川町72番地

㉒ 代 理 人 弁 理 士 鈴 江 武 彦 外3名

明 細 書

1. 発明の名称

キャッシュメモリ装置

2. 特許請求の範囲

(1) プロセッサと主記憶装置との間にキャッシュメモリを設けてなるキャッシュメモリ装置において、

前記プロセッサは、キャッシュメモリをアクセスする命令に伴って、キャッシュメモリの先行ロードを指示する命令と指示しない命令とを持ち、上記先行ロードを指示する命令と共にキャッシュメモリをアクセスする際、先行ロードが指示されていることを前記キャッシュメモリに通知する手段を備え、

前記キャッシュメモリは、先行ロードを指示する命令と共にアクセスされたとき、前記プロセッサがアクセスすると予測されるアドレスのデータを前記主記憶装置からキャッシュメモリに先行ロードすることを特徴とするキャッシュメモリ装置。

(2) キャッシュメモリは、先行ロードを行うか

否かを指示するモード情報を保持する手段を備え、先行ロードを指示する命令と共にアクセスされたとき、上記保持されたモード情報に従って主記憶装置からキャッシュメモリへの先行ロードを行うか否かを決定することを特徴とする請求項(1)に記載のキャッシュメモリ装置。

(3) キャッシュメモリは、先行ロードの対象となるアドレス範囲の情報を保持する手段を備え、先行ロードを指示する命令と共にアクセスされたとき、予測したアドレスが上記保持されたアドレス範囲内にあるときにだけ、主記憶装置からキャッシュメモリへの先行ロードを行うことを特徴とする請求項(1)に記載のキャッシュメモリ装置。

(4) 主記憶装置からキャッシュメモリに先行ロードするデータのアドレスは、プロセッサがアクセスするアドレスに基づいて予測されることを特徴とする請求項(1)に記載のキャッシュメモリ装置。

(5) 先行ロードを指示する命令は、そのオペランドにポインタとオフセット値を持ち、プロセッ

サは上記ポイントにより示されるアドレスにアクセスする際、前記オフセット値をキャッシュメモリに通知する手段を備え、キャッシュメモリは前記命令によりアクセスされたアドレスに前記オフセット値を加えたアドレスを予測アドレスとして先行ロードを行うことを特徴とする請求項(1)に記載のキャッシュメモリ装置。

(6) キャッシュメモリは、メモリ領域を分割してなる複数のウェイを備え、これらのウェイの中の予め定められたウェイに対してのみデータの先行ロードを行うことを特徴とする請求項(1)に記載のキャッシュメモリ装置。

(7) キャッシュメモリは、先行ロードしたデータがアクセスされたか否かの情報を保持する手段をキャッシュブロック毎に持ち、この保持情報に従って先行ロードに伴うデータの追い出しを行うキャッシュブロックを決定することを特徴とする請求項(1)に記載のキャッシュメモリ装置。

さてキャッシュメモリは、基本的にはプロセッサからのメモリアクセスを受け、アクセス要求されたデータがキャッシュデータとして保持されているかを調べた上で、該当データが保持されていないとき、その時点で主記憶装置からキャッシュメモリへのデータ転送を行っている。つまり従来のキャッシュメモリ装置では、オンデマイン드의制御により主記憶装置からキャッシュメモリへのデータ転送を行っている。然し乍ら、このような制御方式は、プロセッサからのメモリアクセスのパターンが局所的である場合には殆ど問題とならないが、メモリアクセスが頻繁にあるような場合にはキャッシュメモリの存在意義が失われると云う不具合がある。

即ち、プロセッサにおけるサイクルタイムの短縮化とアーキテクチャ技術の進歩に伴うプロセッサの処理能力の向上により、今までスーパーコンピュータで処理されていた問題、例えば極めて大きな行列演算処理をマイクロプロセッサを用いて処理することが種々試みられるようになってきた。

3. 発明の詳細な説明

【発明の目的】

(産業上の利用分野)

本発明はプロセッサと主記憶装置との間にキャッシュメモリを設けてメモリアクセスの高速化を図ったキャッシュメモリ装置に関する。

(従来の技術)

プロセッサと主記憶装置との間に設けられて上記主記憶装置に記憶されているデータの一部のコピーを保持する高速・小容量のキャッシュメモリは、上記プロセッサのサイクルタイムが主記憶装置のアクセスタイムに比較して高速である場合、メモリアクセスに対するバッファとして機能して前記プロセッサの見掛上のアクセスタイムの短縮化を図る上で非常に大きな役割を果たす。そして最近では、この種のキャッシュメモリをマイクロプロセッサと共に用いることが常識化しつつあり、VLSI化技術の発展に伴って上述したキャッシュメモリをプロセッサと共にワンチップ化することも種々試みられている。

しかしてこの種のキャッシュメモリ容量よりも大きな配列データをシーケンシャルにアクセスしながら実行される上記行列演算処理を取り扱う場合、そのアクセスのたびにキャッシュミスが発生することが多々生じる。この場合、キャッシュミスが生じる都度、主記憶装置からキャッシュメモリにデータ転送する必要があるため、その間、プロセッサは処理の実行を待たされることになるので、結局、キャッシュメモリの有効性が低下し、プロセッサでの処理が大幅に遅れる、つまりその処理効率が低下すると云う不具合が生じる。

このような不具合を解消するべく、従来、キャッシュミスが発生したとき、アクセスデータを含むデータブロックのみならず、その次のデータブロックのデータまでを一括して主記憶装置からキャッシュメモリに転送し、次のメモリアクセスに対処させようとするのが考えられている。然し乍ら、このような手法を採用すると不必要なデータまで大量にキャッシュメモリに転送することになる上、キャッシュメモリの記憶容量が限られて

いる為、実際に必要とされるデータがキャッシュメモリから不本意に追い出されてしまうと云う不具合が生じ易かった。

(発明が解決しようとする課題)

このように従来のキャッシュメモリ装置においては、プロセッサからのメモリアクセスに対してキャッシュミスが発生した時点でオンデマインドにキャッシュデータの転送を行っているので、行列演算を実行する場合のように或る関連のある一連のデータを頻繁にメモリアクセスするような場合にキャッシュミスが生じ易く、キャッシュメモリを設けた意味合いが失われると云う問題があった。またアクセスデータを含むデータブロックと共に、次のブロックのデータまでを一括してキャッシュ転送するような場合、不必要なデータをキャッシュメモリに転送してしまう上、必要なデータを不本意にキャッシュメモリから追い出してしまうと云う不具合が生じ易かった。

本発明はこのような事情を考慮してなされたもので、その目的とするところは、プロセッサが得

来にアクセスすると予測されるデータを予め効率的にキャッシュメモリに転送しておき、これによって頻繁に生じるメモリアクセスに対するキャッシュミスの発生を抑えることで、プロセッサの処理効率を十分高めることを可能とするキャッシュメモリ装置を提供することにある。

[発明の構成]

(課題を解決するための手段)

本発明に係るキャッシュメモリ装置は、プロセッサにキャッシュメモリをアクセスする命令に伴って、キャッシュメモリの先行ロードを指示する命令と指示しない命令とを準備し、上記先行ロードを指示する命令と共にキャッシュメモリをアクセスする際、先行ロードが指示されていることを前記プロセッサからキャッシュメモリに通知するようにし、

前記キャッシュメモリでは、先行ロードを指示する命令と共にアクセスされたとき、例えば上記先行ロードを指示する命令がアクセスするアドレスに基づいて前記プロセッサが将来アクセスする

と予測されるアドレスを求め、この予測されたアドレスのデータを前記主記憶装置からキャッシュメモリに先行ロードするようにしたことを特徴とするものである。

またキャッシュメモリに、先行ロードを行うか否かを指示するモード情報を保持する手段を設け、先行ロードを指示する命令と共にアクセスされたとき、上記保持されたモード情報に従って主記憶装置からキャッシュメモリへの先行ロードを行うか否かを決定するようにしたことを特徴とするものである。

更にはキャッシュメモリに、先行ロードの対象となるアドレス範囲の情報を保持する手段を設け、予測したアドレスが上記保持されたアドレス範囲内にあるときにだけ、主記憶装置からキャッシュメモリへの先行ロードを行うようにしたことを特徴とするものである。

(作用)

本発明によれば、プロセッサから先行ロードを指示する命令が与えられたとき、キャッシュメ

モリではその命令から予測されるアドレスのデータを主記憶装置から先行ロードするので、例えば大きな配列データをシーケンシャルにアクセスするような場合、プロセッサが次にアクセスすると予想されるアドレスのデータを次々とキャッシュメモリに先行ロードしておくことが可能となる。従ってプロセッサが実際にそのアドレスのデータが必要となった時点には常にそのデータをキャッシュメモリに保持しておくことが可能となる。しかも将来的に必要なと予測されるデータだけを効果的にキャッシュメモリに予め転送しておくことが可能となる。この結果、キャッシュミスの発生を大幅に低減すること、つまりキャッシュメモリに対するヒット率を高めることが可能となり、キャッシュメモリの振能、およびプロセッサの処理能力を十分に発揮させることが可能となる。

(実施例)

以下、図面を参照して本発明の実施例に係るキャッシュメモリ装置について説明する。

第1図は実施例装置の概略構成図で、1は入力

されたプログラムを実行するプロセッサ、2は各種データを格納した主記憶装置（メインメモリ）、3は上記プロセッサ1と主記憶装置2との間に設けられたキャッシュメモリである。このキャッシュメモリ3は高速動作する小容量のメモリ部3aとこのメモリ部3aに対するアクセス動作等を制御するメモリ制御部3bとからなる。このメモリ部3aに前記主記憶装置2に記憶されている各種データの一部が、例えばブロック単位でコピーされて格納される。

この第1図に示す如く構成されるキャッシュメモリ装置は、基本的にはプロセッサ1がプログラムの実行過程においてメモリアクセス（リード／ライト要求）したとき、キャッシュメモリ3にて上記メモリアクセスを受けて動作する。そしてキャッシュメモリ3ではそのメモリ制御部3bの下で上記プロセッサ1がアクセスしたアドレスのデータがメモリ部3aに格納されているか否かを調べ、メモリ部3aに該当データが存在する場合（ヒットした場合）、前記プロセッサ1からのアクセス要

求に応じて該当アドレスのデータの読み出し、または該当アドレスへのデータの書き込みを行う。またメモリ部3aに該当データが存在しない場合（キャッシュミスした場合）には、キャッシュメモリ3は主記憶装置2をアクセスして該当アドレスのデータを、例えばそのデータを含むデータブロックを単位として主記憶装置2からキャッシュメモリ3の前記メモリ部3aに転送した後、前記プロセッサ1からのメモリアクセスに応える。

尚、この主記憶装置2からの新たなデータブロックの転送時には、既にキャッシュメモリ3（メモリ部3a）に格納されているデータの一部が、例えば古いものから順にデータブロック単位で追い出される。

このように構成されたキャッシュメモリ装置にあっては、プロセッサ1からのメモリアクセスに対してキャッシュヒットした場合には、主記憶装置2を一々アクセスするよりも速かに高速にメモリアクセスを実現し、高速にデータのリード／ライトを行うことが可能となる。また逆にキャッ

シュミスした場合には、該当アドレスのデータを主記憶装置2から転送した後にメモリアクセスに応える必要があるため、その間、プロセッサ1での処理実行が待たされることになる。

ここでこの実施例装置が特徴とするところは、プロセッサ1がキャッシュメモリ3をアクセスする命令として、通常のリード／ライト要求を行う命令に加えて、リード／ライト要求と同時にキャッシュメモリ3へのデータの先行ロード指示を行う命令とが準備され、この先行ロードを指示する命令によりキャッシュメモリ3をアクセスする際、プロセッサ1は先行ロードが指示されていることを前記キャッシュメモリ3に対して通知する機能を備えている点にある。

即ち、プロセッサ1における機械語コードを生成するコンパイラは、例えばプログラム中で配列データをシーケンシャルにアクセスする部分が見出されたとき、通常のメモリアクセス命令に代えて、本来のメモリアクセスと同時にキャッシュデータの先行ロードを指示する命令を生成する機能

を備えている。そしてこの先行ロードを指示するメモリアクセス命令を用いてキャッシュメモリ3をアクセスする際、プロセッサ1は先行ロード制御線PLをアクティブにし、その旨（先行ロードを指示する命令であること）をキャッシュメモリ3に通知するものとなっている。

しかしてキャッシュメモリ3では、プロセッサ1からのメモリアクセスを受けたとき、メモリ制御部3bにて前記先行ロード制御線PLがアクティブになっているか否かを調べる。そして上述した先行ロードが指示されていない常のリード／ライト命令である場合には、従来装置と同様にキャッシュ動作し、キャッシュミスした場合には該当アドレスのデータを主記憶装置2からキャッシュメモリ3に転送して前記メモリアクセスに応える。

これに対してプロセッサ1から与えられたアクセス要求が前記先行ロードを指示するリード／ライト命令である場合には、キャッシュメモリ3のメモリ制御部3bは、先ずその本来の命令であるリード／ライト要求に従って従来装置と同様にキャ

ッシュ動作する。そしてこの本来のメモリアクセスに対するキャッシュ動作を実行した後、或いはこのメモリアクセス動作と並行して前記プロセッサ1から与えられた命令に基づいて前記プロセッサ1が将来アクセスすると予測されるアドレスのデータを前記主記憶装置2から先行ロードする。この主記憶装置2からのデータの先行ロードは、例えば予測されたアドレスのデータを含むデータブロックのデータを1単位として前記メモリ部3aに転送することによりなされる。

このデータの先行ロードについて更に詳しく説明すると、例えばキャッシュメモリ8のメモリ制御部3bにはデータブロックの先行ロードを制御する為の第2図に示すようなレジスタが設けられている。このレジスタは、先行ロードの指示をインヒビットするビットフィールドINH、およびデータブロックの先行ロードを許可するアドレス範囲（先頭アドレスと最終アドレス）を特定するアドレスフィールドを持つ。尚、この第2図では1つのアドレス範囲を指定する先頭アドレスと最終

アドレスのフィールドだけが示されるが、アドレス範囲としては2つ以上分散して指定するようにすることもできる。この場合には、第1の先頭アドレスと第1の最終アドレス、第2の先頭アドレスと第2の最終アドレス、……と云うように、個々のアドレス範囲をそれぞれ指定し得るようなフィールド構成としておけば良い。

しかし最も単純な先行ロードすべきアドレスの予測は、例えばメモリアクセス命令によってアクセスされたアドレスのデータが含まれるデータブロックの次のデータブロックを指定するアドレスを予測値とする方式である。

ここで従来考えられていたデータの先行ロードと本質的に異なる点は、従来装置にあっては、上述したような先行ロードを指示する命令と先行ロードを指示しない命令との2種類がなく、一般的なメモリアクセス命令（本発明における先行ロードを指示しない命令）が与えられたときに無条件にそのアクセスされたアドレスのデータを含むデータブロックの次のデータブロックを先行ロード

しているのに対し、本発明では上記先行ロードを指示する命令が与えられたときにだけ先行ロードの制御を行う点にある。つまりプロセッサ1からキャッシュメモリ8に対して先行ロードを指示するメモリアクセス命令を与えるか、或いは先行ロードを指示しないメモリアクセス命令を与えるかによってキャッシュメモリ8における先行ロードを制御するようにしている点で従来考えられていた先行ロードの方式とは本質的に異なっている。

しかもこのような先行ロードを指示する命令がプロセッサ1から与えられたとき、キャッシュメモリ8では、前述したレジスタに保持されている先行ロード制御の為の情報をを用いて、指示された先行ロードを実行するか否かをキャッシュメモリ8自体で制御している点を従来考えられていた先行ロードの方式と全く異にしている。

即ち、この実施例装置では、先行ロードを指示する命令が与えられたとき、先ず前記レジスタのビットフィールドINHのインヒビット情報に従い、上記先行ロードを指示する命令をインヒビットす

る。そしてその命令がインヒビットされないとき、次に前記先行ロードを許可するアドレス範囲の情報に従い、先行ロードしようとするデータ（データブロック）のアドレスが許可範囲内にあるか否かを調べる。このようにしてアクセス要求のあった命令から予測される先行ロードの対象となるアドレスがレジスタに保持された先行ロード許可アドレス範囲内にあることが確認されたとき、メモリ制御部3bは主記憶装置2をアクセスし、その予測されたアドレスのデータを含むデータブロックをキャッシュメモリ8に転送し、前記メモリ部3aに格納する。このような予測アドレスに基づく主記憶装置2からのデータブロックのキャッシュメモリ8への転送により、プロセッサ1が将来アクセスすると予測されるアドレスのデータが先行ロードされることになる。

尚、このようなキャッシュメモリ8への主記憶装置2からのデータの先行ロードは、基本的にはプロセッサ1における命令の実行とは独立に、換言すればプロセッサ1での命令実行と並行して行

われる。然し乍ら、キャッシュメモリ3における先行ロードの実行中にプロセッサ1から次のメモリアクセス命令が与えられ、そのメモリアクセスがキャッシュミスしたような場合には、例えば上記先行ロードの処理手続きを中断し、キャッシュミスに伴う主記憶装置2からのキャッシュメモリ3へのデータ転送を優先的に処理するようにすることが望ましい。そしてキャッシュメモリ3としては、多少キャッシュメモリ3としての性能が低下するが、前記プロセッサ1がキャッシュメモリ3に対してアクセスを行っていない期間を利用して上述した先行ロードを実行するようにすれば良い。

しかしキャッシュメモリ3にアクセスポートを2つ設けておき、プロセッサ1がキャッシュメモリ3をアクセス中でも、先に予測されたアドレスのデータブロックの先行ロードを上記アクセスに並行して行い得るようにしておけば、プロセッサ1の処理動作(メモリアクセス動作)とは全く独立にメモリアクセス命令に基づいて予測したアドレ

うにしても良い。

また或いは、メモリアクセスに伴ってデータの先行ロードを行う命令を複数種類設けておき、前述した先行ロード制御線PLを利用してその命令の種類をキャッシュメモリ3に通知するようにする。そしてキャッシュメモリ3では、命令の種類毎にその最後のアクセスアドレスを保持しておき、これらの各命令の種類毎に新たにアクセスされたアドレスとその命令について保持されている最後のアクセスアドレスとの差分を求め、この差分を新たなアクセスアドレスに加えて予測アドレスを求めるようにしても良い。

つまり先にアクセスされたアドレスと、現在アクセスされたアドレスとの差分を現アクセスアドレスに加えたアドレスを予測値としてデータの先行ロードを行うようにしても良い。このようにしてプロセッサ1が将来アクセスすると予測されるアドレスを予測する場合であっても、前述したようにインヒビットや先行ロード許可領域のアドレス範囲に対する判定を行った上で先行ロードを行

スのデータブロックの先行ロードを行うことが可能となる。

但し、上述したようにキャッシュメモリ3が2つのアクセスポートを備えている場合であっても、例えばデータブロックの先行ロードを実行している期間に新たなキャッシュミスが発生し、これによってデータブロックの入れ替えが必要となるような場合には、一旦、上記データブロックの先行ロードを待機させ、上記データブロックの入れ替えが終了した後、データブロックの先行ロードを再開することが必要があることは言うまでもない。

ところで先行ロードするデータのアドレスを予測する方式として、前述したようにアクセス要求のあったアドレスのデータを含むデータブロックの次のデータブロックのアドレスを予測値として求めることのみならず、例えばキャッシュメモリ3にてページ単位で最後にアクセスしたアドレスを保持しておき、このアドレスと新たにアクセスされたアドレスとの差分を上記新たなアクセスアドレスに加えて、これを予測アドレス値とするよ

えば良いことは勿論のことである。

かくしてこのように構成された本実施例装置によれば、例えば第3図にキャッシュメモリ3からプロセッサ1に次々とデータをロードする場合のシステム的な動作の流れを示すように、先行ロードを指示する命令を用いてキャッシュメモリ3をアクセスすると、その指示に従ってアクセスされた命令に基づいて予測されたアドレスのデータが次々と先行ロードされる。具体的には、A番地をアクセスし、キャッシュミスした場合にはA番地のデータを主記憶装置2から転送し上記アクセス要求に応えた後、上記アクセスされたA番地から予測されるB番地のデータの先行ロードが行われる。その後、先行ロードを指示する命令を用いてB番地がアクセスされると、このB番地のデータが先行ロードされていることからキャッシュヒットし、B番地のデータがキャッシュメモリ3からプロセッサ1に直接的に読み込まれる。このとき、キャッシュメモリ3は上記B番地から次にアクセスされると予測されるC番地のデータを主記憶装

置2から先行ロードする。このような先行ロードを指示する命令に従い、次々と先行ロードすべきデータのアドレスを予測し、その予測アドレスのデータを先行ロードしておくことにより、実際にプロセッサ1が当該アドレスのデータをアクセスしたとき、そのアドレスのデータをキャッシュメモリ3に保持しておくことが可能となるので、キャッシュミス招くことなく高速にメモリアクセスに対処することが可能となる。

尚、先行ロードを指示する命令を用いてメモリアクセスがなされたとき、第3図に示すようにD番地がアクセスされ、このD番地から予測されたアドレスが前述した先行ロード許可領域のアドレス範囲外となった場合には、当該予測アドレスの先行ロードは行われない。

ところでキャッシュメモリ3に予測アドレスのデータを先行ロードする場合、何等かの形で既にキャッシュメモリ3に格納されているデータの一部を追い出すことが必要である。このデータの追い出しは基本的には、古いデータから順に追い出

すことによりなされるが、単純に古いデータから順に追い出すと不具合が発生することがある。

例えば通常の配列計算を行うプログラムでは、配列データに対する演算を行うループ内であっても、配列データのみならずそれ以外の変数データに対するアクセスが多々行われる。また上記ループ以外では配列データ以外のデータに対するアクセスが殆どである。従って前述したデータブロックの先行ロードを大規模な配列データに対して行った場合、上述した配列データ以外のデータを不本意に次々と追い出してしまおうと云う不具合が生じることが予想される。

このような不具合に対処するには、例えばキャッシュメモリ3（メモリ部3a）のメモリ空間を複数の領域に分割して複数のウェイを構成し、これらのウェイ毎にデータの先行ロードを制御するようにすれば良い。具体的には、例えばウェイ[0]からウェイ[3]の4つのウェイにキャッシュメモリ空間を分割し、その中のウェイ[0]からウェイ[2]だけを先行ロードの対象として制限を

加えるようにする。このようにして前述した配列演算に用いられる配列データとそれ以外のデータとがそれぞれ格納されるキャッシュメモリ3におけるウェイを可能な限り分離する。そして前記配列データはシーケンシャルにアクセスされることが多いと云う性質、換言すれば1度アクセスされた配列データは、その後、暫くの間はアクセスされることがないと云う性質に従って、当該配列データに対するアクセスにおいてキャッシュミスが発生した場合や、その配列データを先行ロードするような場合、その配列に属する別のデータブロック、特に一旦アクセスされたことのあるデータブロックをキャッシュメモリ3から追い出すようにすれば良い。

このようにして先行ロードの対象となるウェイを限定して予測アドレスのデータの先行ロードを行うようにすれば、データの先行ロードにより効果が期待される配列データ以外のデータについては、そのままキャッシュメモリ3に残したまま、配列データの間でだけデータの追い出しと先行ロ

ードとを実行することが可能となる。つまり必要なデータを不本意にキャッシュメモリ3から追い出すことなしに、プロセッサ1が将来アクセスすると予測されるアドレスのデータ（データブロック）を効果的に先行ロードすることが可能となる。

尚、キャッシュメモリ3内に設定するウェイの数、またその中の先行ロードを許可するウェイの数とそのウェイ番号はシステム仕様に応じて任意に定めれば良いものである。またデータブロックの先行ロードを行うウェイを、例えば先行ロードの指示がなされた命令の本来のメモリアクセスに対してヒットしたウェイとして定めたり、或いは本来のメモリアクセスがキャッシュミスし、そのデータブロックがロードされたウェイとして定めることも可能である。更には上述した如く特定されるウェイの選択優先度を高め、他のウェイとの兼ね合いに応じて先行ロードするウェイを決定するようにすることも可能である。

さて本発明は次のようにして実施することも可能である。前述した実施例では、プロセッサ1が

ら先行ロードを指示する命令が与えられたとき、先行ロード制御線 P_Lを用いてキャッシュメモリ 3 にその旨を通知するようにしたが、将来アクセスすると予測されるデータブロックのアドレス予測の為の付加情報（予測情報）を前記アクセス命令と共にキャッシュメモリに与えるようにしても良い。

例えば第4図に示すように命令コードと共に、その命令のオペランド情報としてソースレジスタ番号、ポインタ、オフセット値を予測情報としてキャッシュメモリ 3 に与えるようにする。この命令フォーマットは、例えばポインタの値をアドレスとしてメモリアクセスを行い、且つ同時に上記ポインタの値をオフセット値だけ順次インクリメントする命令機能を兼ね備える。

このような機能を持つ命令は、データブロックの先行ロード機能を備えていなくても有効な命令であり、或る意味では多くの計算機で一般的にサポートされている命令形態である。しかしてこのような命令は、スタックアクセスや配列アクセス

が多く、インクリメントされたポインタの値に対して近いアドレスを、将来プロセッサ1がアクセスする確率が高い場合に多く用いられる。従ってこのような命令を用いて先行ロードを指示し、プロセッサ1によりインクリメントされたポインタの値を予測アドレスとすれば、先行ロードするデータのアドレスとしての的中率は極めて高くなる。

尚、このような命令に付属した予測情報を用いる場合には、例えば命令のポインタの値により示されるアドレスにて本来のメモリアクセスを行い、このメモリアクセスアドレスに前記オフセット値を加えたアドレスを予測アドレスとして先行ロードを行うことのみならず、ポインタの値にオフセット値を加えたアドレスにアクセスし、その上でそのアクセスアドレスに前記オフセット値を更に加えたアドレスを予測アドレスとして先行ロードを行うようにしても良い。そして上記オフセット値をキャッシュメモリ 3 に保持し、プロセッサ1では前記アクセスアドレスに前記ポインタ値を加えたアドレスを次の命令のアクセスアドレスとし

て順次設定していくようにすることも可能である。

このようなアドレス制御を行えば、先行ロードを指示する命令をキャッシュメモリ 3 に与えるだけで、プロセッサ1では次々とメモリアクセスのアドレス指定を行い、キャッシュメモリ 3 ではそのメモリアクセスに先立って該当アドレスのデータを次々に先行ロードしていくことが可能となる。

従って、上述したようにデータブロックの先行ロードを指示する命令自体を、従来から用いられているポインタ、オフセット値付きの命令機能を拡張したものとして用いることにより、従来のオブジェクトプログラムをそのまま用いて前述した先行ロード機能を実現することができる。従って従来システムとのソフトウェアの互換性を保つ上で非常に有用であると云える。

尚、ソフトウェアの互換性がさほど問題とならないような場合には、前述した命令とは別に先行ロードを行う為のアドレス情報を生成するためのオペランドを準備しておくようにすれば良い。

かくして上述した本発明の実施例に示されるよ

うに、本発明によればコンパイラがデータブロックの先行ロードを実行したほうがメモリアクセスに対する効果が高いと判断した場合にのみ、プロセッサ1からキャッシュメモリ 3 に対して先行ロードを指示する命令を与え、この先行ロードを指示する命令が与えられた場合にだけキャッシュメモリ 3 は本来のメモリアクセスのアドレスに基づいて予測されるアドレスのデータを先行ロードする。

従って先行ロードしたデータに対するメモリアクセスのヒット率を十分に高めることが可能となる。しかもプロセッサ1が近い将来アクセスすると予測されるアドレスのデータを含むデータブロックだけを、例えばキャッシュメモリ 3 のクエイを制限しながらメモリアクセスをトリガとして先行ロードするので、必要なデータを不本意に追い出すことなしに、必要なデータ、或いは必要となると予測されるデータだけをキャッシュすることができる。故にアクセスアドレスの予測を、例えば前述したポインタ値等に基づいて効果的に行え

ば、キャッシュミスの発生を殆ど零に抑え、キャッシュメモリ3の効果をも十分に引き出してプロセッサ1での処理効率を効果的に高めることが可能となる。

ちなみに従来のように、メモリアクセスがあったとき、そのアクセスされたアドレスのデータブロックと共に次のデータブロックまでをその都度一括してキャッシュメモリに転送する方式では、不必要なデータや遠い将来にしか必要としないデータまでが先行ロードされてしまうと云う問題がある。つまり従来の先行ロード方式は、単にアクセスされたデータを基準としてキャッシュメモリに転送するデータの幅を広げているに過ぎないと云える。この結果、不必要なデータの先行ロードに伴い、キャッシュメモリ3から必要なデータブロックが追い出されてしまうことが否めず、キャッシュに対するヒット率の向上が期待できないと云う問題がある。しかもデータブロックの先行ロードがキャッシュミスをトリガとして行われるだけなので、例えば一次元の配列データを単位シーケ

ンシャルにアクセスするような場合、2データブロックに1回しかその先行ロードが行われないことになる。このことは、キャッシュのヒット率の向上をさほど望めないことを意味する。

この点、本発明にあっては、配列データをシーケンシャルにアクセスするような場合にだけ先行ロードを指示する命令を用いてメモリアクセスし、そのメモリアクセスがヒットするかキャッシュミスするかには拘らず、アクセスされた命令に基づいて予測されるアドレスのデータを先行ロードするので、そのヒット率を効果的に高めることができると云う実用上多大なる効果が奏せられる。

尚、本発明は上述した実施例に限定されるものではない。例えばキャッシュメモリ3に設定するインヒビットの情報は、キャッシュメモリ3の内部状態等に応じて設定すれば良いものである。また先行ロードを許可するアドレス範囲についても、例えば主記憶装置2に記憶されているデータのメモリマップ情報等に応じて定めれば良い。

またキャッシュブロック単位に先行ロードされ

たキャッシュブロックが実際にアクセスされたか否かを示すフラグを設け、或るデータブロックの先行ロードを行う為に別のデータブロックを追いつく必要がある場合、上記フラグ情報に従ってデータの追いつきを行うデータブロックを決定するようにすることも可能である。また同時にフラグの足っていないウェイの追いつき優先度を高めて、どのウェイからデータブロックの追いつきを行うかを決定するようにすることも可能である。

更にはメモリアクセスされたアドレスに基づく先行ロードの対象とするアドレスの予測についても、前述したポインタとオフセット値を用いることのみならず、例えば配列データの構造に基づいて予測決定するようにすることも可能である。その他、本発明はその要旨を逸脱しない範囲で種々変形して実施することができる。

〔発明の効果〕

以上説明したように本発明によれば、メモリアクセスするデータの内容に応じて、先行ロードを指示する命令と先行ロードを指示しない命令と

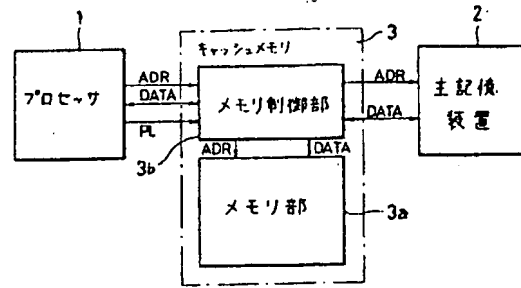
を使い分け、例えば配列データをシーケンシャルにアクセスするような場合にのみ、プロセッサが近い将来アクセスすると予測されるアドレスのデータをキャッシュメモリに先行ロードする。しかもプロセッサからの先行ロードを指示する命令が与えられることをトリガとして、そのメモリアクセスがヒットするかミスするかには拘らず、予測されたアドレスのデータを先行ロードするものとなっている。この結果、大量の配列データをシーケンシャルにアクセスするような場合であって、シーケンシャルにアクセスされるデータを次々とキャッシュメモリに先行ロードしておくことが可能となるので、キャッシュミスの発生を抑えてプロセッサの処理効率を効果的に高めることが可能となる等の実用上多大なる効果が奏せられる。

4. 図面の簡単な説明

図は本発明の一実施例に係るキャッシュメモリ装置について示すもので、第1図は実施例装置の概略構成図、第2図は実施例装置のキャッシュメモリに設けられる先行ロード制御の為のレジス

タ格納情報を示す図、第3図は実施例装置の効果
を説明するためのシステムの処理の流れを示す
図、第4図は先行ロードするデータブロックのア
ドレスを予測する為の予測情報の構成例を示す図
である。

1…プロセッサ、2…主記憶装置、3…
キャッシュメモリ、3a…メモリ部、3b…メモリ
制御部。



第1図

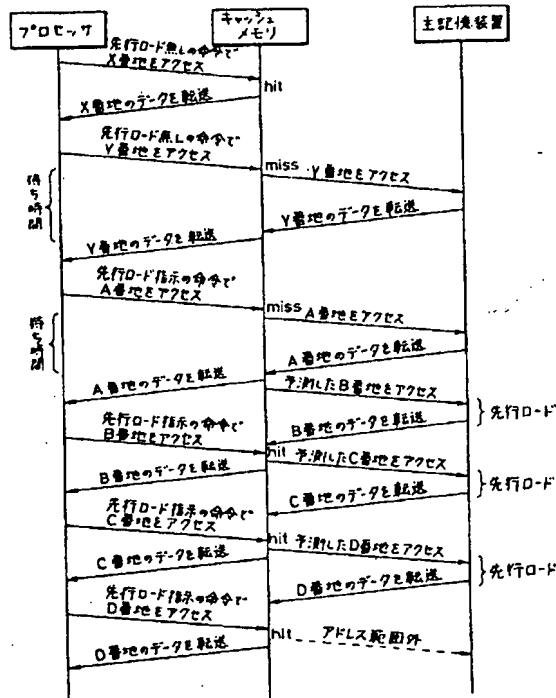
出願人代理人 井理士 鈴江 武彦

INH	先行ロード許可領域 先頭アドレス	先行ロード許可領域 最終アドレス
-----	---------------------	---------------------

第2図

命令コード	ソースレジスタ 番号	ポインタ	オフセット
-------	---------------	------	-------

第4図



第3図

特開平4-52741

【公報種別】特許法第17条の2の規定による補正の掲載
【部門区分】第6部門第3区分
【発行日】平成10年(1998)12月18日

【公開番号】特開平4-52741
【公開日】平成4年(1992)2月20日
【年通号数】公開特許公報4-528
【出願番号】特願平2-155776
【国際特許分類第6版】

G06F 12/08
12/12

【F I】

G06F 12/08 M
12/12 A

予 照 査 正 補

平成 9 年 12 月 18 日

特許庁長官 荒井 尚光 殿

1. 事件の表示

特願平2-155776号

2. 発明の名称

キャッシュメモリ装置およびプロセッサ

3. 補正をする者

事件との関係 特許出願人

(207) 株式会社 東 芝

4. 代理人

東京都千代田区豊田3丁目7番2号

特許内閣特許事務所内

〒100 電話03(3502)3181(大代表)

(5847) 弁護士 神 江 武 彦

5. 目 録 補 正

6. 補正により減少する請求項の数 4

7. 補正対象事項名

(1) 明細書

8. 補正対象項目名

- (1) 発明の名称
- (2) 特許請求の範囲
- (3) 発明の詳細な説明

9. 補正の内容

- (1) 発明の名称を「キャッシュメモリ装置およびプロセッサ」と訂正する。
- (2) 特許請求の範囲の記載を訂正の通り訂正する。
- (3) 明細書第4頁第5行の「キャッシュメモリ装置」を「キャッシュメモリ装置およびキャッシュメモリ装置にデータの転送を要求するプロセッサ」と訂正する。
- (4) 明細書第8頁第5行～第6行の「キャッシュメモリ装置を提供することにある。」を「キャッシュメモリ装置およびキャッシュメモリ装置にデータの転送を要求するプロセッサを提供することにある。」と訂正する。
- (5) 明細書第9頁第17行の後に下記の記載を挿入する。

記

また、本発明はキャッシュメモリ装置に対してデータの転送を要求するプロセッサにおいて、キャッシュメモリ装置のキャッシュメモリをアクセスする命令とともに、主記憶装置から該キャッシュメモリに所定のデータを先行ロードする命令を実行することを特徴とするものである。

2. 物性素の構成

(1) プロセッサと主記憶装置との間に設けられるキャッシュメモリ装置において、

前記プロセッサからの先行ロードを指示する命令により先行ロードが指示された場合に、前記プロセッサがアクセスすると予測されるアドレスのデータを前記主記憶装置からの先行ロードによりキャッシュメモリに保持することを特徴とするキャッシュメモリ装置。

(2) プロセッサと主記憶装置との間に設けられるキャッシュメモリ装置において、

先行ロードの対象となるアドレス範囲の情報を保持する手段と、
前記プロセッサからの先行ロードを指示する命令により先行ロードが指示された場合に、前記プロセッサがアクセスするアドレスを予測する手段と、

この手段により予測されたアドレスが前記アドレス範囲内にある場合に、前記データを前記主記憶装置からの先行ロードによりキャッシュメモリに保持する手段とを備えたことを特徴とするキャッシュメモリ装置。

(3) キャッシュメモリ装置に対してデータの転送を要求するプロセッサにおいて、

前記キャッシュメモリ装置のキャッシュメモリをアクセスする命令とともに、主記憶装置から該キャッシュメモリに所定のデータを先行ロードする命令を実行することを特徴とするプロセッサ。

出願人代理人 弁理士 船江武彦